

# Lexicalization, Separation and transitivity: A comparative study of Mandarin VO compound Variations

Menghan Jiang and Chu-Ren Huang

Department of CBS, The Hong Kong Polytechnic University

menghan.jiang@connect.polyu.hk;

churen.huang@polyu.edu.hk

## Abstract

Our study takes a comparable corpus-based statistical approach, to empirically examine the correlation between transitivity and separation ability for VO compound in Mandarin Chinese. The results of the two studies show that inseparable VOs are more likely to be used in a transitive way, compared to separable ones. In addition, there is a statistical negative correlation between transitivity and separation ability, i.e. the more a VO sequence is lexicalized, the less likely it can take an object. Our paper further empirically proves that the grammatical variations of VO compound are to a large extent depend on the degree of lexicalization. The differences in separation and transitivity between Mainland and Taiwan actually indicate the different stages that Mainland and Taiwan VO compounds are located in the continuum of lexicalization.

## 1 Introduction

In Modern Chinese, there is an increasing number of disyllabic VO compounds which gradually changed from intransitive to transitive verbs. The transitive VO compounds can take another constituent (e.g., a word, a phrase or a sentence) as their objects, and yield the configuration of [VO<sub>1</sub>+O<sub>2</sub>], such as 投资 房地产 *touzi fnagdichan* 'throw\_money\_real-estate' 'invest in real estate', 进军 美国 市场 *jinjun meiguo shichang* 'march\_towards\_American\_market' 'march towards American market'. This phenomenon has attracted the interests of numbers of scholars in Chinese linguistics (e.g., Liu, 1998a, 1998b; Gao, 1998 among others). One research question that often being addressed is the transition requirement of VO compounds (i.e. what kind of VO is easier to be transferred from intransitive to transitive). Numbers of researches claim that for a VO compound, the ability of taking the object is closely related to its lexical status. The higher

degree of lexical status, the more possibility it can take the object and be used transitively (e.g., Liu, 1998a; Luo, 1998; Gao, 1998). Actually, this is in accordance with Brinton and Traugott (2005) which claims that lexicalization is to use a syntactic construction or word formation as a new form, which cannot be completely derivable or predictable from the constituents of the construction or the word formation pattern.

It is also well known that the degree of lexicalization can be tested through separation test (e.g., Her, 1997; Liu, 1998a). The easier it can be separated, the higher degree of its lexicalization. In fact, this is related to the 'Lexical Integrity Hypothesis' proposed by Huang (1984: 60): no phrase-level rule may affect a proper sub-part of a word. Since a VO compound as a word is thus a lexical unit whose internal structure is of a V+O (Her, 1997), and an important feature that distinguishes a lexical units from a phrase is the lexical integrity.

Therefore based on the previous discussions, it has become a common belief among linguistic researchers that there is a strong correlation between the transitivity of VO and whether the VO is separable (the lexical status), i.e. the VO which cannot be separated is much more likely to be used as a transitive verb, and vice versa. For example, Gao (1998) has classified VO into three types according to their separation ability: VO can be separated without constraints (e.g., 着急 *zhaoji* 'worry', 放心 *fangxin* 'reassure', 发愁 *fachou* 'be anxious'), VO can be separated with constraints (e.g., 毕业 *biye* 'graduate', 担心 *danxin* 'anxious', 留心 *liuxin* 'be careful', 害怕 *haipa* 'be scare') and VO cannot be separated (e.g., 出版 *chuban* 'publish', 当心 *dangxin* 'take care', 动员 *dongyuan* 'mobilize'). After investigating some of the VO in the corpus, he then concludes that all the VOs that cannot be separated are used as a

transitive verb (e.g., 动员 群众 *dongyuan qunzhong* ‘mobilize the masses’) while the VOs which can be separated without constraints are usually cannot be used transitively (e.g., \*放心 他的 能力 *fangxin ta de nengli* put\_heart\_he\_DE<sup>1</sup>\_ability ‘rest assured his ability’). For the VOs that can be separated with constraints, they usually have transitive usages in the corpus (e.g., 担心 工程 的 进度 *danxin gongcheng de jindu* worry\_about\_project\_DE\_progress ‘worry about the progress of the project’), but some of the words are still under the process of changing (e.g., ?过目 这 份 文件 *guomu zhefen wenjian* look over\_this\_CL<sup>2</sup>\_document ‘look over this document’).

One thing should be noted is although the correlation between transitivity and separation for a VO has been well recognized by linguists, in literature we can barely find empirical study using real data to verify this common belief. For the very few studies (e.g., the study of Gao (1998) we mentioned above) that are conducted based on empirical data, their data size is relatively small and the statistical methods they are using are also quite simple (often just percentage or pure numbers). Although the numbers and percentages can reveal the difference, they cannot tell whether there is significance or not.

Therefore it is important for us to investigate this issue in a more empirical and quantitative way, with the assistance of large-scale corpus as well as the statistical tool. In that sense, the correlation between transitivity and lexical status can be verified systematically and comprehensively.

Another point often ignored by previous researches is that, although there are numbers of researches discussing the transitivity and separation ability of VO compounds, the variation difference between different variants of the same language are lack of studied. There are a very few study using relatively small set of data to point out that Taiwan and Singapore VO compounds have higher transitivity frequency (e.g., Wang, 1997; Diao, 1998) and Mainland words tend to have more separation usages than Taiwan (Diao, 2016). But the relationship of transitivity and separation

<sup>1</sup>的 DE: particle which appears between the modifier and the head noun

<sup>2</sup> CL: classifier.

between language variations has not been examined. Then we would also like to ask questions: are there any transitivity differences between Mainland and Taiwan Mandarin? If the variation difference in transitivity exists, is this variation dependent on the degree of lexicalization? In other words, whether the variation differences in transitivity indicate the different stages that VO compounds from different variants are located in the continuum/process of lexicalization?

## 2 Data collection and calculation

### 2.1 Measurement of separation ability

Therefore, our first aim is to examine the relationship between transitivity frequency and lexical status of VO sequences, with the assistance of large-scale comparable corpus. It should be noted that previous studies usually examine both separation status and transitivity issue in a dichotomy way. In other words, the VO is classified as separable vs. inseparable, transitive vs. intransitive (e.g., Gao, 1998; Her, 1996 among others). But we argue that the issues of both separation and transitivity are not simply binary dichotomy, it is more about tendency/frequency difference. For example, both 把关 *baguan* guard\_pass ‘guarantee’ and 插手 *chashou* ‘intervene’ are separable (e.g., 把 了 关 *ba le guan* insert\_hand ‘guaranteed’; 插 过 手 *cha guo shou* ‘have intervened’), but the frequency of separation usages are very different (把关 *baguan* ‘guarantee a pass’ is much more frequently to be used separately than 插手 *chashou* ‘intervene’). In addition, the grammatical elements which can be inserted also vary a lot for these two words. Plenty of elements can be inserted into 把关 *baguan* (把 产 品 质 量 关 *ba chanpin zhiliang guan* guard\_product\_quality\_pass ‘guarantee the quality of products’; 把 好 了 进 出 口 检 验 关 *ba hao le jin chukou jianyan guan* guard\_good\_LE<sup>3</sup> import\_export\_inspect\_pass ‘have guaranteed the inspection of import and export’) while only aspectual marker can insert into 插手 *chashou* ‘intervene’ (插 过 手 / 插 了 手 *cha guo shou/cha le*

<sup>3</sup>了 LE: perfective marker.

*shou* ‘have intervened/intervened’). In terms of transitivity of VO compound, the transitivity degree also varies a lot. For example, although both 驰名 *chiming* ‘famous’ and 约会 *yuehui* ‘date’ can be used transitively, the frequency of using as a transitive verb for 驰名 *chiming* (e.g., 驰名中外 *chiming zhongwai* ‘renowned both inside and outside the country’) is much higher than that of 约会 *yuehui* (e.g., 约会拜金女 *yuehui baijinnü* ‘date material girl’). In that sense, we argue in our paper that examining the transitivity and separation issue in a continuous way would reflect the real situation of language more objectively.

In this study, we use frequency/percentage of separation usages to measure the separation degree.

**relative frequency=separated usages/all the usages** (e.g., Ren and Wang, 2005)

Example: separation frequency for 操心 *caoxin* ‘worry about’ = the number of 操...心 usages (10 tokens)/all the usages of 操心 *caoxin* (287 tokens) + Separation usages (10 tokens) = 3.367%

## 2.2 Measurement of transitivity degree

The transitivity of VOs is measured by frequency also: **transitivity frequency=transitive tokens/all the tokens**. For example, transitivity frequency of 签约 *qianyue* ‘sign a contract’ =number of transitive usages of 签约 *qianyue* (13 tokens)/all the tokens of 签约 *qianyue* (1000 tokens) = 1.3%. The VO word list we use in this study is the same as we have used in the previous section: the 109 VO compounds which we have collected from previous researches (e.g., Qian, 2011; Luo, 1998). But in this study, we exclude 13 words that do not show significant variation difference in both transitivity frequency and Chi-square test: 登场 *dengchang* ‘show’、操心 *caoxin* ‘worry about’、致信 *zhixin* ‘write letter to’、出土 *chutu* ‘be unearthed’、参演 *canyan* ‘act in’、更名 *gengming* ‘rename’、涉嫌 *shexian* ‘be suspected’、领军 *lingjun* ‘play a leading role’、揭秘 *jiemi* ‘expose’、解码 *jiema* ‘decoding’、启航 *qihang* ‘set sail’、失信 *shixin* ‘break promise’、移情 *yiqing* ‘love someone else’. For all the 96 words in our wordlist, we calculate their separation and

transitivity frequency in both Mainland and Taiwan.

## 3 Data analysis and result

### 3.1 Study 1: Comparison between separable words and inseparable words

The 96 words in the wordlist are divided into two categories according to their separation frequency: the VOs that have separable usages (separation frequency >0) in the corpus (here we call it “separable VO compound”) and the VOs that do not have any separate usages (separation frequency =0) in the corpus (“inseparable VO compounds”)<sup>4</sup>. Based on their separation status, we ask the first research question: is there a significant difference in transitivity frequency between these two groups (Empirically and statistically, is it true that the inseparable VOs are more likely to be used in a transitive way, and vice versa)? The data distribution of separable and inseparable words in both Mainland and Taiwan varieties is shown in table 1.

	Taiwan	Mainland
Separable VO	39	37
Inseparable VO	57	59

Table 1. Number of separable/inseparable VO

As we can see from the table, the numbers of separable VO and inseparable VO are close and the distributions in Mainland and Taiwan are also very similar, which make the comparison more reasonable.

A Mann-Whitney U test was run to determine if there were differences in transitivity frequency between separable and inseparable VO compounds in each variety. Mann-Whitney U test is often presented as the non-parametric alternative to independent-sample t-test, as it does not require the normality of the data, it is very suitable for our current study. The statistical tool we use is IBM SPSS V.22.

The result of Mann Whitney U test for Taiwan data is shown in Table2. It displays that the median value of transitivity frequency for inseparable VO compounds (0.3607142857) is significantly higher

<sup>4</sup> For the ‘inseparable VO’, we are not claiming that separation is impossible under any context. But since Gigaword corpus is very large, if no separation usages are detected in the corpus, the separation frequency should be very low.

than that for separable VO compounds (median value is 0.1378091873),  $U=801.000$ ,  $Z=-2.316$ ,  $P=0.021$ .

Hypothesis Test Summary

Null Hypothesis	Test	Sig.	Decision
The distribution of TW transitivity is the same across categories of separation type	Independent-Samples Mann-Whitney U Test	0.021	Reject the null hypothesis

Asymptotic significances are displayed. The significance level is .05

Total N	96
Mann-Whitney U	801.000
Wilcoxon W	1581.000
Test Statistic	801.000
Standard Error	134.049
Standardized Test Statistic	-2.316
Asymptotic Sig. (2-sided test)	.021

Table 2. Mann Whitney U test for TW data

Mainland data presents the same result, table 3 shows that median value of transitivity frequency for inseparable compounds in Mainland is 0.278, which is statistically significant higher than that of separable VO compounds (0.076),  $U=761.500$ ,  $Z=-2.485$ ,  $P=0.013$ .

Hypothesis Test Summary

Null Hypothesis	Test	Sig.	Decision
The distribution of ML transitivity is the same across categories of separation type	Independent-Samples Mann-Whitney U Test	0.013	Reject the null hypothesis

Asymptotic significances are displayed. The significance level is .05

Total N	96
Mann-Whitney U	761.500
Wilcoxon W	1464.500
Test Statistic	761.500
Standard Error	132.811
Standardized Test Statistic	-2.485
Asymptotic Sig. (2-sided test)	.013

Table 3. Mann Whitney U test for Mainland data

Summary for study 1: The results of Mann Whitney U test in both Mainland and Taiwan Mandarin show that the transitivity frequencies between separable and inseparable VO compounds are significantly different in both varieties. In other words, in both Mainland and Taiwan Mandarin, we can observe empirically that inseparable VOs are much more likely to be used in a transitive way.

### 3.2 Study 2: Correlation between separation ability and transitivity

In the first study, we have shown that compared to separable VO compounds, the inseparable ones are more likely to be used as a transitive verb. But as we have mentioned in the first section, the separation frequencies vary a lot among separable VO compounds. Therefore, what would be the case if we consider the separation frequency: Is there a significant statistical correlation between the separation frequency and the transitivity frequency of the VOs? (i.e. is it empirically true that the more frequently it is used separately, the less frequent it can be used transitively?)

In this study, the separation frequency (separation frequency=separated usages/all the usages) is included as a variable for statistical analysis. We use the Spearman's rank-order correlation to assess the relationship between transitivity frequency and separation frequency in both Taiwan and Mainland Mandarin. The result of Taiwan data is shown in the table.

Correlations

			TW separation fre	TW Transitivity
Spearman's rho	TW separation fre	Correlation coefficient Sig. (2-tailed) N	1.000 . 96	-.221* 0.030 96
	TW transitivity	Correlation coefficient Sig. (2-tailed) N	-.221* 0.030 96	1.000 . 96

\*.Correlation is significant at the 0.05 level (2-tailed).

Table 4. Spearman's correlation for TW data

The result indicates that there exist a negative correlation between transitivity frequency and separation frequency in Mainland Mandarin, the correlation is statistically significant,  $r=-0.221$ ,  $P=0.03$ .

We have the similar result for Mainland data, as shown below. There is a significant negative correlation between transitivity frequency and separation frequency in Taiwan Mandarin:  $r=-0.237$ ;  $P=0.02$ .

Correlations

			ML separation fre	ML Transitivity
Spearman's rho	ML separation fre	Correlation coefficient Sig. (2-tailed) N	1.000 . 96	-.237* 0.020 96
	ML transitivity	Correlation coefficient Sig. (2-tailed) N	-.221* 0.030 96	1.000 . 96

\*.Correlation is significant at the 0.05 level (2-tailed).

Table 5. Spearman's correlation for Mainland data

The results of spearman's correlation in both Taiwan and Mainland show that for a VO compound, the more frequently it is used separately, the less likely it can be used as a transitive verb. In other words, for a VO compound, the more it is lexicalized, the more likely it is used in a transitive way.

Summary for study 2, based on the result of the two empirical studies we have conducted, the tendency can be observed is that, compared to separable VO compounds, inseparable ones are more likely to be used in a transitive way. And also, for a VO compound, the less frequently it is used separately, the more likely it is used as a transitive verb. In other words, if a VO sequence is less lexicalized, its probability of being transitive is higher. The tendency is in accordance with what has been presented in the previous papers and is true for both Mainland and Taiwan data.

But it should also be noted that although the result of our second study show that there is a significant negative correlation between transitivity and separation, the correlation coefficients in both Mainland and Taiwan are to some extent low, which indicates that the negative correlation is relatively weak in both varieties. This can be explained because although the lexical status of a VO compound does affect the transitivity, it is not the only factor. In the real language, there are varieties of factors which are influencing the transitivity apart from the lexical status. The factors include not only some internal linguistic factors (e.g., word frequency; the degree of freedom for each morpheme; event type of the verb), but also some external social factors (e.g., the influence of social media or other languages/dialects).

#### 4 Grammatical Variation and Lexicalization

As we discussed in the above section, the transitivity of a VO compound is statistically significant correlated with separation ability (which is measured by separation frequency) in both Mainland and Taiwan Mandarin. Then one question needed to be asked is: are there any variation differences in transitivity between Mainland and Taiwan Mandarin? If the answer is yes, does this transitivity difference is depend on the process of lexicalization of these VO compounds? In other words, do the differences of transitivity and separation between Taiwan and Mainland Mandarin indicate the different stages that Mainland and Taiwan VO compounds are located in the continuum/process of lexicalization?

In order to have a general picture of the data distribution, first we start from the comparison between average transitivity and separation frequency in Mainland and Taiwan Mandarin. As the table below displays, the average transitivity frequency of Taiwan VO compounds (0.3538) is higher than that of Mainland Mandarin (0.2919) whereas the separation frequency of Taiwan VO compounds (0.00707) is obviously lower than that of Mainland VO compounds.

	Taiwan VO	Mainland VO
Average transitivity fre	0.3538	0.2919
Average separation fre	0.007068073	0.019565008

Table 6. Average transitivity and separation

The first impression is that in general, the Taiwan VO compounds are more likely to have transitive usages while their Mainland counterparts have more probabilities to be used separately. But the average can only give us a general tendency about data distribution, and more statistical tests (e.g., Z-test, likelihood ratio test) are still needed to carefully examine the variation difference in transitivity frequency and separation frequency. We will illustrate the statistical analysis in detail in the following section.

In terms of the transitivity frequency, Z-test is conducted to investigate whether the transitivity frequencies between these two varieties have significant differences. According to the result of

Z-test (shown in the figure below), among all the 96 words we include in our study, 76 VO compounds show significant differences in transitivity frequency between the two varieties while 20 words are not significant different. Among the 76 words, 53 Taiwan VO compounds show significant higher transitivity frequency than their Mainland counterparts and 23 VOs have significantly higher transitivity frequency in Mainland than in Taiwan usages. In this sense, we can see the clear tendency that Taiwan VO compounds tend to be more likely to have higher transitivity usages.

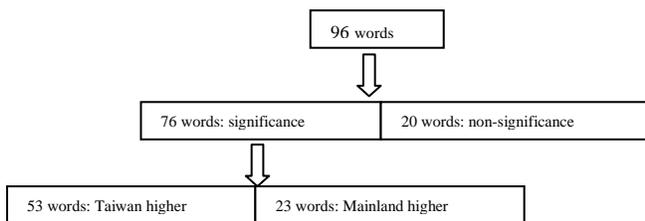


Figure 1. The result of Z-test

The VO compounds can be categorized into three types according to the Z-test result: the VO whose transitivity frequency in Taiwan is significantly higher than in Mainland (Taiwan transitivity higher); the VO whose transitivity frequency in Mainland is significantly higher than in Taiwan (Mainland transitivity higher); and there is no significant difference in transitivity frequency between Mainland and Taiwan (no transitivity difference). Based on this classification, we found that for the compounds in “Taiwan transitivity higher” group, their separation frequencies in Taiwan are much lower. For the other two groups (“Mainland transitivity higher” and “no transitivity difference”), the differences in separation frequency between varieties are not very obvious.

Transitivity frequency	Mainland separation frequency	TW separation frequency
TW transitivity higher	0.0579	0.015
ML transitivity higher	0.0251	0.019
No significance	0.008	0.002

Table 7. separation difference based on Z-test

So far, the general tendency is clear: the transitivity of Taiwan VO compound is

significantly higher, especially for the words whose Taiwan transitivity is significantly higher than their Mainland counterparts. But one thing should be noted is that the P value cannot tell us everything. In other words, among the 76 words which show significant difference in transitivity frequency between Mainland and Taiwan, their degree of difference varies. For example (as shown in table 8), the transitivity frequencies of both 过境 *guojing* ‘transit boarder’ and 借道 *jiedao* ‘channeled through’ have significant difference between Mainland and Taiwan at  $P < 0.01$  level, but for 借道 *jiedao*, its transitivity frequencies in Taiwan and Mainland are quite close (0.871383/0.689655) whereas the transitivity of 过境 *guojing* in two varieties actually have much bigger difference (0.341/0.033). To solve the problem, likelihood ratio test is also used in our study to measure the degree of variation difference. The formulation is shown below: likelihood ratio = higher frequency/lower frequency. For 过境 *guojing*, the likelihood ratio of Taiwan to Mainland is 10.33 ( $=0.341/0.033$ ), meaning that Taiwan 过境 *guojing* is about 10 times more likely to be used as a transitive verb than the Mainland counterpart while for 借道 *jiedao*, the likelihood ratio of Taiwan to Mainland is only 1.26, which is much lower than the one of 过境 *guojing*. And this actually indicates that the transitivity difference of 借道 *jiedao* between Mainland and Taiwan is not as obvious as that of 过境 *guojing*.

	P value for Z-test	TW Transitivity	Transitivity in Mainland	Likelihood ratio
过境	<0.01	0.341	0.033	10.33
借道	<0.01	0.871	0.690	1.26

Table 8. Comparison between Z-test and likelihood

We calculate the likelihood ratio variation for all the 76 words which show significance in Z-test. Based on the result of likelihood ratio test, the tendency difference between Mainland and Taiwan becomes clearer. When the likelihood ratio is larger than 10 (ratio  $\geq 10$ ), we consider the two varieties to have prominent significant differences in transitivity frequency. And we found for the 8 words which belong to this group, all of them have higher transitivity in Taiwan, in other words, our

data shows that Taiwan VO compounds have significantly higher transitivity, especially when the difference is prominent. And when the transitivity of Taiwan VO compounds is prominent higher than that of Mainland VO, the separation frequency between the two varieties are also observed to have prominent difference: but this time, the separation frequency of Mainland is significantly higher than Taiwan.

For example, the transitivity frequency of 把关 in Taiwan (24.5%) is significantly higher than in Mainland (0.71%).

	TW	ML
Separation usages	43 types, 59 tokens	906 types, 1808 tokens
Separation frequency	1.19%	45.75%
Transitivity frequency	24.5%	0.71%
Examples	<p>把好质量关 <i>bahao zhiliang guan</i> guard_good_quality_pass 'guarantee the quality'</p>	<p>天津市 严把进津企业资质审验关 <i>tianjinshi yanba jinjin qite zizhi shenyan guan</i> Tianjin_strict_guard_enter_Tianjin_enterprise_qualification_verification_pass 'Tianjin strictly guarantee the enterprise qualification'</p> <p>把好建设前期工作质量关 <i>bahao jianshe qianqi gongzuo zhiliang guan</i> guard_good_construction_preparatory_work_quality_pass 'guarantee the quality of preparatory work of construction'</p> <p>把了/过关 <i>ba le/guo guan</i> guard_LE/GUO<sup>5</sup>_pass 'guaranteed/have guaranteed'</p>

Table 9. separation comparison of 把关

<sup>5</sup>过 GUO: experiential marker.

The differences in separation usages between Mainland and Taiwan are also very obvious (not only in separation frequency, but also in the grammatical elements can be inserted). The separation frequency of Mainland 把关 *baguan* (45.74%) is significantly higher than that of Taiwan counterpart (1.19%), with a likelihood ratio of 38.437, indicating that 把关 *baguan* is about 38 times more likely to be used separately in Mainland than in Taiwan. Furthermore, the corpus data shows that very few grammatical elements can be inserted into Taiwan 把关 *baguan* (examples like 严把质量关 *yan ba zhiliang guan* 'strictly check the quality' is frequently appeared in Taiwan corpus) while varieties of elements can be inserted into 把关 *baguan* in Mainland Mandarin (e.g., aspectual marker 把了/过关 *ba le/guo guan* 'checked/have checked'; classifier 把好几道关 *ba haojido guan* 'carefully check for several times'; the object 把质量关 *ba zhiliang guan* 'guarantee the quality'; and even the object with modifier 把好进津企业资质审验关 *ba hao jin jin qiye zizhi shenyan guan* 'Tianjin strictly guarantee the enterprise qualification', etc.).

Moreover, the words which have prominent significant variation differences in transitivity frequency are also observed to have contrast differences in separation frequency<sup>6</sup> (i.e. separation usages can only be detected in Mainland corpus). Examples are shown below:

VO	ML examples	TW
撤军 <i>chejun</i> 'withdraw troop'	从约旦河撤了军 <i>cong yuedanhe che le jun</i> from_Jordan_River_Withdrawal_LE_troops 'pull troops out of the River Jordan'	Not detected
联手 <i>lianshou</i> 'join hands'	需要香港和内地联手 <i>xuyao xianggang he neidi lianqishou</i> need_Hong_Kong_and_Mainland_join_up_hand 'need the alliance between Hong Kong and Mainland'	Not detected
献计	为改革发展献一计 <i>wei</i>	Not detected

<sup>6</sup> Although no separation example was found in the corpus, we are not claiming that there is no separation usage in other context. But we argue that since the Gigaword corpus is very large (contains more than 1.1 billion characters), if no separation example was detected in the corpus, the separation frequency should be very low.

<i>xianji</i> 'offer advice'	<i>gaige fazhan xianyiji</i> for_reformation_development _offer_one_advice 'offer advice to reformation development'	
移民 yimin 'immigrate'	移了民 <i>yi le min</i> move_LE_nationality 'have immigrated'	Not detected

Table 10. Contrast difference between ML and TW

To summarize what we have found so far, the separation frequencies of Mainland VO compounds are obviously higher than that of Taiwan VO compounds. For the VOs whose transitivity frequencies in Taiwan are significant higher than in Mainland, their separation frequencies in Taiwan are significantly lower, especially when two varieties have prominent significant differences in transitivity frequency, their differences in separation frequency are also prominent, sometimes even have contrast difference (separation usages can only be detected in Mainland corpus).

Therefore it is possible for us to argue that the differences in transitivity frequency and separation frequency between Mainland and Taiwan Mandarin actually indicate the different stages that Mainland and Taiwan VO compounds are located in the continuum/process of lexicalization. In particular, if the status of lexicalization is considered as a continuum from phrase to word, then compared to Mainland VO compounds, Taiwan VOs behave more like words instead of phrases, therefore it is more likely for the Taiwan VO sequences to be used in a transitive way.

But as we have pointed in section 3, the lexical status is not the only factor which can affect the transitivity of VO compound. A Variety of factors (both internal linguistic factor and external social factor) are also influencing the degree of transitivity. Therefore the variation difference between transitivity and lexical status is not absolute. Exceptions always exist. For example, 感恩 *ganen* 'be thankful' in Mainland has a relatively high separation frequency (0.067797) while in Taiwan the separation frequency is 0.005. Mainland 感恩 *ganen* is about 380 times more likely to have separate usages than its Taiwan counterpart. But the transitivity difference between the two varieties for 感恩 *ganen* is not significant.

It may imply that other factors are actually influencing both transitivity and separation ability. Therefore what we report here is a general tendency of two variants, and the significance of statistical results indicates that the tendency we have proposed is reliable and convincing.

## 5 Conclusion

In our study, we take a large corpus-based statistical approach to examine the correlation between separation and transitivity of VO compound. The results prove that empirically compared to separable VO compounds, inseparable ones are more likely to be used in a transitive way. And also, for a VO compound, the less frequently it is used separately, the more likely it is used as a transitive verb. In other words, if a VO sequence is less lexicalized, its probability of taking an object is higher. But it should be noted that separation ability is not the only factor that is affecting the transitivity of a VO compound, therefore the correlation coefficient of statistical analysis is not very high. In terms of grammatical variation between Taiwan and Mainland Mandarin, our paper further compare the transitivity of VO compound between the two varieties and argue that the differences in separation and transitivity between Mainland and Taiwan actually indicate the different stages that Mainland and Taiwan VO compounds are located in the continuum of lexicalization.

## References

- Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press.
- Her, O. S. (1997). *Interaction and variation in the Chinese VO construction*. Crane Publishing Company.
- Huang, C. T. J. (1984). Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association*, 19(2), 53-78.
- Diao, Y. B. (1998). Study on VO+O construction. 也谈“动宾式动词+ 宾语”

- 形式. *Chinese Construction*. *语文建设*, (6), 39-41.
- Diao, Y. B. (2016). The comparisons of separable words usages cross-strait. 海峡两岸离合词使用情况对比考察. *Education of overseas Chinese*, *海外华文教育*, (4), 435-446.
- Gao, G. (1998). Collation rules of VO+O. “动宾式动词+ 宾语” 的搭配规律. *Chinese Construction*. *语文建设*, (6), 36-38.
- Liu, D. W. (1998a) Thoughts on VO+O construction (1). 关于动宾带宾现象的一些思考 (上). *Chinese Construction*. *语文建设*, (1), 22-26.
- Liu, D. W. (1998b). Thoughts on VO+O construction (2). 关于动宾带宾现象的一些思考 (下). *Chinese Construction*. *语文建设*, (3), 28-29.
- Luo, X. R. (1998). Exploration on VO+O. *Chinese Construction*. *语文建设*, (5), 27-30.
- Qian, C. Y. (2011). *Analysis on VO+O construction in Modern Chinese*. 现代汉语“动宾式复合词带宾语” 结构分析. MA thesis. Fudan University.
- H. D. Wang. (1997). What is the rule/pattern for VO+O? “动宾式动词+ 宾语” 规律何在?. *语文建设*, (8), 30-31.